

A New Signal Feature Extraction Method - Fractal dimensions of Time-Frequency Domain

Yuan Yu Wang TingYu Li Baoliang Yao Shixuan

School of mechanical engineering, Dalian JiaoTong University, Dalian, China, 116028

Yuanyu_knife@163.com

Abstract

In this paper, vibration signal multifractal method was researched, multifractal on time-frequency domain was an energy fractal and the signal feature extraction was based on the analysis of its energy distributing. The method analysed signal on time-frequency domain to characterize the distributing of its frequency or energy, and the signal's feature was extracted by fractal dimension. After the signal was changed to time-frequency domain by Hilbert-Huang transform, general dimension D_q would be calculated from the signal in time-frequency domain by least square method. In the end, examples of emulator and practical application proved that this integrated method was feasible.

Key words: multifractal; Hilbert-Huang transform; general dimension

1. Introduction

For many faults, there will be some complex vibration phenomenon sometimes, because the fault system is a nonlinear system. Measured vibration signals often mixed the periodic motion, the non-periodic motion, the system noise, the measurement noise and so on. When we use the traditional theory method to describe the complexity, we find that the theory itself becomes more and more limited. In recent years, using the fractal theory of the non-linear science to describe and research the vibration form of mechanical systems has the preliminary application, but these applications are used by a single fractal theory. In practical applications, the single fractal theory only reflects the scrambling of a signal in whole, and doesn't adequately describe the characteristics of the local singularity. The research shows that the multifractal can in extenso describe the distribution form of the fractal singular chance of the signal, and can increase the fine-scale of describing geometrical features and the behavior of the local-scale of the signal. [1, 2, 3]

The scholars at home and abroad have already done a lot of research work for applying the fractal theory to the vibration signal feature extraction. But the known studies are mostly confined to the fractal analysis of the vibration signal in the time domain. For the non-linear and non-stationary signals, the analysis is not enough only in the time domain [4]. The traditional spectral analysis only works under the stable conditions hypothesis; the DEA (Data Envelopment Analysis) needs the detailed criteria to distinguish between the natural and random fluctuation and the fluctuation led to the mechanical failure; the JTFA (Joint Time-Frequency Analysis)[4] which synthesized the merit of both the spectral analysis and the DEA more suits the analysis of this kind of non-linear and non-steady signals such as the reciprocating mechanical vibration signal. In this paper, a method of a vibration signal multifractal in the time-frequency domain is presented. This method realizes the signal analysis, which is based on the quantification feature extraction of the energy distribution of the signal in the time-frequency domain.

2. Overview of the multifractal

Multifractal means that if the bounded set in the d dimensional Euclidean space is the carried set of a random variable, the corresponding probability distribution has the singular measure, i.e. the set as the bounded set is made up of an infinite amount of subsets by overlap, and moreover, the probability measure has the singular feature which changes with the scale index point by point, with the inexistence of the corresponding density function. The multifractal also is called the fractal measure. The invariant measures of non-linear dynamic systems in various fields often belong to the fractal measure. The multifractal theory quantificationally describes the distribution of the fractal measure in the entire carried set. [5, 6]

The following equation is the definition of generalized dimension, which Renyi first introduced

and B.B.Mandelbrot rediscovered in the study on the singular set.

$$D_q = -\lim_{\varepsilon \rightarrow 0} \lg K_q(\varepsilon) / \lg \varepsilon \quad (1)$$

In the equation,

$$K_q(\varepsilon) = \left\{ \log \sum_{i=1}^N (P_i)^q \right\} / (1-q) \quad (2)$$

K_q means the generalized entropy, and q means the scale index. Thus, with the changes of q we can get,

When $q=0$, D_q is the box dimension,

$$D_0 = \frac{\lim_{\varepsilon \rightarrow 0} \lg K_0}{\lg \frac{1}{\varepsilon}} \quad (3)$$

When $q=1$, D_q is the information dimension,

$$D_1 = \lim_{\varepsilon \rightarrow 0} \sum_i \frac{P_i \lg P_i}{\lg \varepsilon} \quad (4)$$

When $q=2$, D_q is the correlation dimension,

$$D_2 = \lim_{\varepsilon \rightarrow 0} \lg \sum_i \frac{P_i^2}{\lg \varepsilon} \quad (5)$$

In that way, D_3, D_4, D_5 and others can be got.

3. The numerical algorithm of the generalized dimension in the time-frequency domain

3.1 To Disperse The Signal

We suppose that there is a existing signal $X(t)$, T is sampling length and Δt is sampling period. Thus the sampling points $l=T/\Delta t$, the signal is dispersed to time series $X(i)$, $i=1,2,3,\dots,l$.

3.2 Gridding Partition of Sampling Space

As shown in Fig1, taking ε_j as grid width, thus the number of rows and columns of the grid is $F_j = T/\varepsilon_j$, $j=1,2,3,\dots,J$. We suppose that the m th row and the n th column is mn grid, J is the kind of the gridding partition, so the coordinate of gridding of mn grid is shown as Fig 2 as followed:

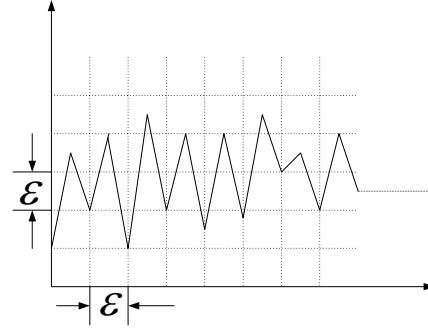


Fig1 Grid Division of Sample Space

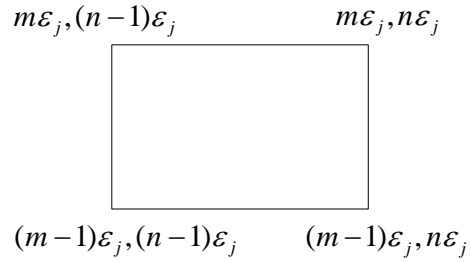


Fig 2 Coordinate of gridding
Gridding partition of sampling space

We mark the points of mn grid which covers collection as d_{mn} , so the probability that the mn grid covers aggregate is

$$P_{mn}(\varepsilon_j) = \frac{d_{mn}}{\sum_{m=1}^S \sum_{n=1}^S d_{mn}} = d_{mn} / l$$

Thus from function (1) and (2), we can get a set of K_q .

3.3 The Calculation of Generalized Dimension in time domain

We suppose that $T(j) = \log(\varepsilon_j)$, $S_j = K_q(\varepsilon_j)$

$$S(j) = -D_q T(j) + B$$

and D_q is the slope of a straight line is information dimension, B is intercept, thus we can establish the function $f(D_q, B) = \sum_{j=1}^J [S(j) + D_q T(j) - B]^2$, from the

least square method, we can see the condition which can make this function get its minimum is

$$\begin{cases} \frac{\partial f}{\partial D_q} = 2 \sum_{j=1}^J [S(j) + D_q T(j) - B] [T(j)] = 0 \\ \frac{\partial f}{\partial B} = -2 \sum_{j=1}^J [S(j) + D_q T(j) - B] = 0 \end{cases}$$

So :

$$\begin{aligned} D_q(\varepsilon_j) &= \frac{J \sum_{j=1}^J T(j) S(j) - \sum_{j=1}^J S(j) \sum_{j=1}^J T(j)}{J \sum_{j=1}^J T(j)^2 - [\sum_{j=1}^J T(j)]^2} \\ &= \frac{J \sum_{j=1}^J K_q(\varepsilon_j) \lg \varepsilon(j) - \sum_{j=1}^J \lg \varepsilon(j) \sum_{j=1}^J K_q(\varepsilon_j)}{J \sum_{j=1}^J (K_q(\varepsilon_j))^2 - [\sum_{j=1}^J K_q(\varepsilon_j)]^2} \end{aligned}$$

We call the maximum of D_q obtained from the function as sensitive dimension in time domain.

4. The numerical algorithm of the generalized dimension in the time-frequency domain

4.1 The time-frequency decomposition and normalized total energy of the signal

X_n is the known time-series. When X_n is changed by Hilbert-Huang transform, $J(\omega, t) = \sum_{i=1}^n b_i A_i(t) e^{j\omega_i(t)}$

is got, and the obtained time-frequency images are showed as the gray-scale images. After the normalizing, the amplitude matrix A_{UV} can be obtained. Each spot in the matrix corresponds to a time and instantaneous frequency. The numerical square of the spot expresses the normalized energy size at this time and instantaneous frequency in the time domain. So the normalized total energy of the time-frequency plane is

$$E = \sum_{u=1}^U \sum_{v=1}^V A_{uv}^2.$$

4.2 Gridding partition of time-frequency space

Being similar to the partition of the time-domain signal sample space, the gridding width is ε_j . J is the type number of the gridding partition. The number of the gridding lines and lists is $R_j = F / \varepsilon_j$ and $P_j = T / \varepsilon_j$. T is the sampling length, and F is the analysis frequency. Supposing that the m th line and

n th row gridding is the mn gridding, and J is the type number of the gridding partition, the coordinate chart of the mn grid is shown in Figure 3.

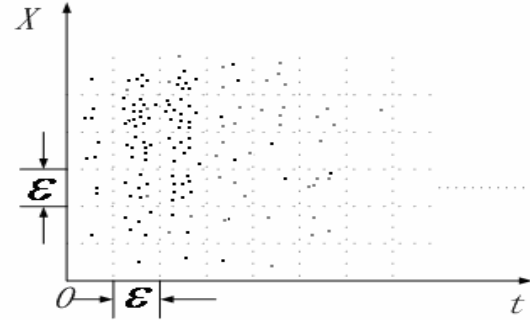


Fig 3. Gridding partition of sampling space

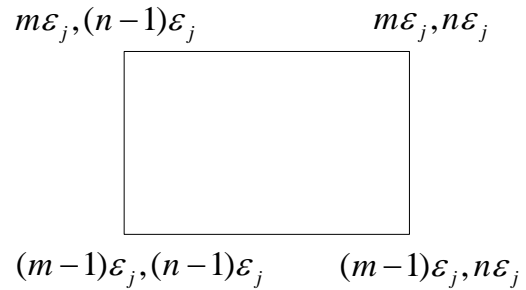


Fig 4. Coordinate of gridding

Let E_{mn} be an aggregate of the energy covered with the gridding. Then the probability of an aggregate of the gridding coverage is,

$$P = (\varepsilon_j) = \frac{E_{mn}}{\sum_{m=1}^R \sum_{n=1}^P E_{mn}} = E_{mn} / E \quad (6)$$

A series of $K_q(\varepsilon_j)$ can be obtained by Formula 3

3.3 The Calculation of Generalized Dimension in time-frequency domain

We suppose that $T(j) = \log(\varepsilon_j)$, $S_j = K_q(\varepsilon_j)$

$$S(j) = -D_q T(j) + B$$

and D_q is the slope of a straight line is information dimension, B is intercept, thus we can establish the function $f(D_q, B) = \sum_{j=1}^J [S(j) + D_q T(j) - B]^2$, from the least square method, we can see the condition which can make this function get its minimum is

$$\begin{cases} \frac{\partial f}{\partial D_q} = 2 \sum_{j=1}^J [S(j) + D_q T(j) - B][T(j)] = 0 \\ \frac{\partial f}{\partial D_q} = -2 \sum_{j=1}^J [S(j) + D_q T(j) - B] = 0 \end{cases}$$

So :

$$D_q(\varepsilon_j) = \frac{J \sum_{j=1}^J T(j)S(j) - \sum_{j=1}^J S(j) \sum_{j=1}^J T(j)}{J \sum_{j=1}^J T(j)^2 - [\sum_{j=1}^J T(j)]^2}$$

$$= \frac{J \sum_{j=1}^J K_q(\varepsilon_j) \lg \varepsilon(j) - \sum_{j=1}^J \lg \varepsilon(j) \sum_{j=1}^J K_q(\varepsilon_j)}{J \sum_{j=1}^J (K_q(\varepsilon_j))^2 - [\sum_{j=1}^J K_q(\varepsilon_j)]^2}$$

We call the maximum of D_q obtained from the function as sensitive dimension in time-frequency domain.

5. The analysis of emulational examples

We use the program named atoms which produces a basic Gauss Yuan linear combination signal in Matlab to produces two group of Gauss Yuan linear combination signal. The signal 1 is atoms (128,[45,0.25,32,1;85,0.25,32,1]), the signal 2 is atoms (128,[45,0.15,32,1;85,0.32,32,1]). The time-domain waveforms and time-frequency distributions of the two groups of signals are shown in the following figures.

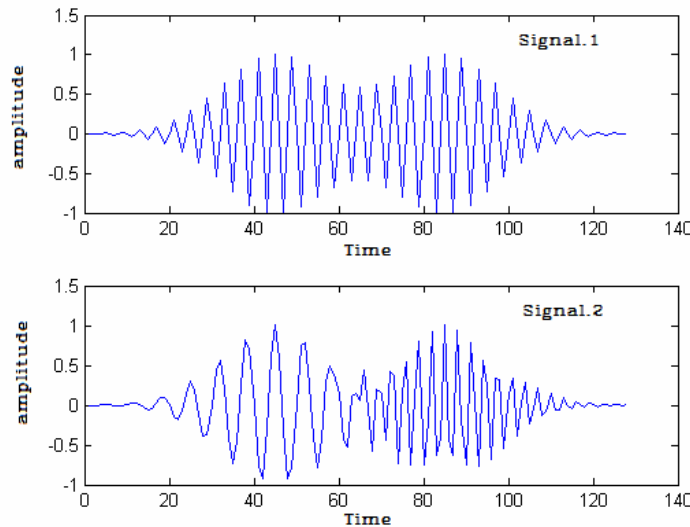


Fig 5. Emulational signal

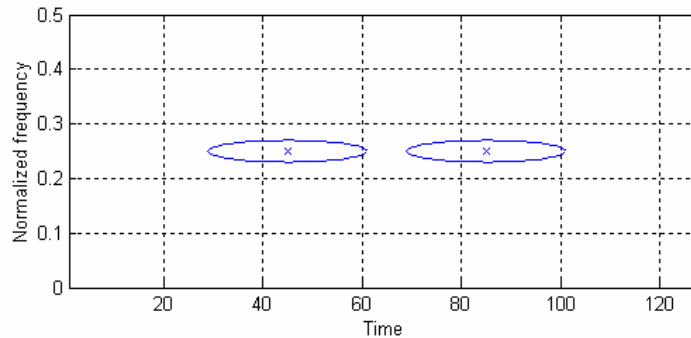


Fig 6. Time-Frequency distribution of emulational signal1

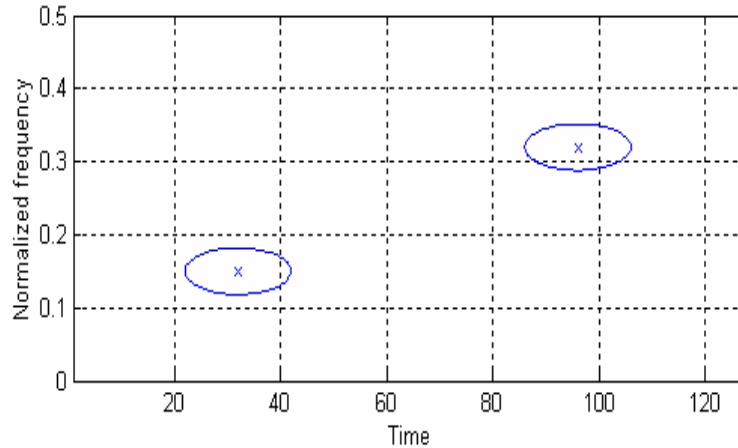


Figure 7. Time-Frequency distribution of emulational signal2

By figure5, 6, and 7, the differences between the time-domain waveforms of emulational signal1 and 2 are not very obvious, but their time-frequency domains obviously reflect the differences between the two signals. There are obvious differences between the gaussian kernel distributions of the two signals.

The result is as follows, which is got by the multifractal in time domain and time-frequency domain of two signals.

Tab. 1 Multifractal in time domain of two types of vibration signals

	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
Signal 1	1.894	1.249	0.604	0.512	0.177	0.273	0.332	0.373	0.404	0.429
Signal 2	1.894	1.265	0.636	0.549	0.137	0.229	0.282	0.314	0.334	0.347

Tab. 2 Multifractal in time-frequency domain of two types of vibration signals

	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
Signal 1	1.879	0.940	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Signal 2	1.879	1.227	0.574	0.447	0.399	0.374	0.359	0.349	0.342	0.336

There is not big difference between the two emulational signals in time domain fractal, which can't reflect the similarities and differences between the two signals. The similarities and differences between the two signals can be reflected by the fractal in time-frequency domain, so the fractal in time-frequency domain integrates more information, which is more conducive to the feature extraction.

polyethylene production line will be out of production, and the losses after that will be immeasurable. In order to predict the equipment operational state, as well as prevent the equipment from fatal accidents, it is necessary to diagnose C-1150 Ultra-high pressure reciprocating compressor preliminarily. Fig8 shows the compressor.

6. Project example analysis

6.1 Engineering Background

C-1150 Ultra high pressure reciprocating compressor is a pivotal equipment of a certain chemical plant. Most high-pressure ethylene gas come from the plant is compressed by the equipment mentioned above. If C-1150 broke down, the whole

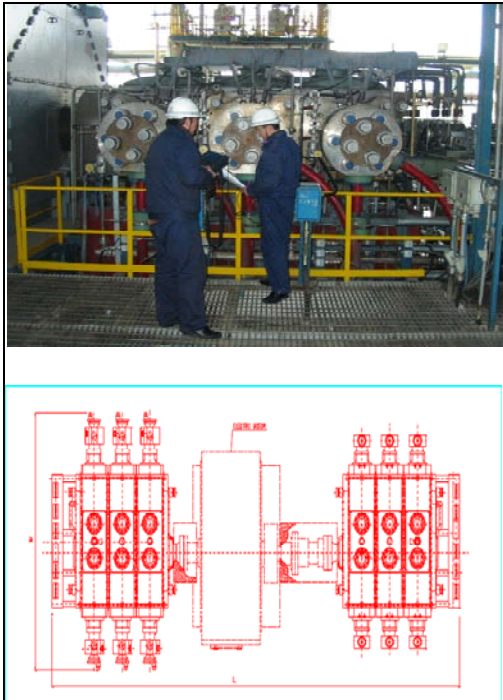


Fig 8 C-1150 compressor

Concrete information for C-1150 is as followed:
Tab.3 C-1150 Information

Parameter Name	Parameter index
Compressor Crank Number	6
Compressor Cylinder Number	12
Compressor Forms	Contraposed
Compressor Serial Number	12PK/2
Driver signature (KW/	16500/214
Admission Pressure	About 25MPa
discharge pressure	About 280MPa

6.2 Measuring Points Collocation

(1) Parameter setting: Sample length : 16k, Sample frequency: 25.6kHz, Sensor sensitivity: 9.837mv/G.

(2) Diagram of Measuring Points:

Diagram of measuring Points is shown as Fig.9:

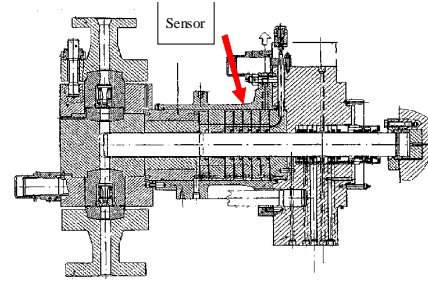


Fig.9 Structure of packing

6.3 Measuring Point Analysis

During the monitoring process to C-1150 Ultra-high pressure reciprocating compressor, leakage in various degree happened at the packing of 6 # cylinder. We will analyze 3 groups of vibration signals near the packing contrastively. Signal explication is shown as Tab.4:

Tab.4 Contrast of three types of vibration signal of packing

Fault	Acquisitio	T (°C)	Signal Peak
Normal	2004.6.10	33	18.33
Slight	2004.7.4	-1	24.02
Serious	2004.11.22	-15	21.1

From Tab.4 we can see that the degree of packing leakage is closely related with the temperature of the inlet pipe behind the cylinder packing which has been monitored by the system, when the leakage happened, volume expands sharply owing to the high pressure gas coming out of cylinder. The process of high pressure gas expansion is endothermic process, thus the temperature of the inlet pipe will be lower and even frost if getting serious. The peak of vibration signal reaches its maximum in the early state, the value in the serious state is lower than that in the early state, because the peak group of the vibration signal near the packing is impacted by high-pressure airflow when turning on or turning off the valve. In the early state, a little gas in the cylinder comes out from the gap between packing ring and piston, and this will lead to the increase of high-frequency components of vibration signal, the vibration energy and peak value will increase as well. However, the amount of gas leakage in the serious state is more than that in the normal state, gaseous tension get lower, the impact happened when turning on or turning off the valve will be slighter than that in the normal state, so the peak will get lower in a certain degree.

Time-domain waveform of the 3 kinds of vibration signal is shown as Fig10:

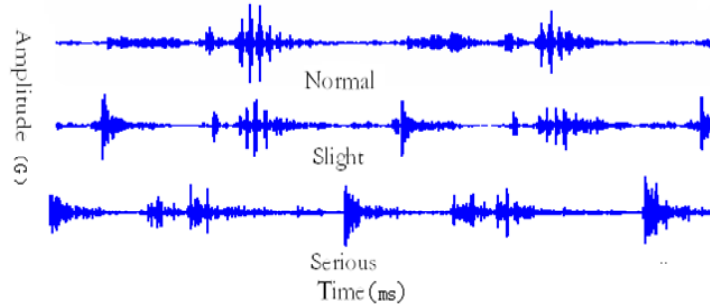


Figure 10. Time domain wave of three types of vibration signals

The following calculated the multifractal of the vibration signal in time domain and the multifractal of the local wave in time-frequency domain, compared and analysed them.

After getting the multifractal of the three groups of signals in a work cycle, the general dimension of the signal in time domain is shown in the TABLE.5.(Only a part of the data is listed, because of the length limits.)

Tab.5 Multifractal contrast in time domain

	D_0	D_1	D_2	D_3	D_4	D_5
normal	2.9091	1.4978	1.2310	0.3577	0.0781	0.0505
Slight divulging	2.8791	2.1115	2.4828	1.2777	0.9007	0.7273
Serious divulging	2.9294	1.4768	1.4789	0.5006	0.1686	0.0628

Tab.6 Multifractal contrast in Time-frequency domain

	Df_0	Df_1	Df_2	Df_3	Df_4	Df_5
normal	1.9880	0.8718	0.3524	0.3562	0.3608	0.3600
Slight divulging	1.9880	1.7323	0.5304	0.5921	0.6181	0.6316
Serious divulging	1.9883	2.0566	0.6544	0.6797	0.6913	0.6913

From Table.5 we can see that when q equals other value except 1, the dimension in the early state is higher than that in the normal and serious state, and the value in the normal state is similar to that in the serious state. So the complexity of the time domain vibration signal in the early state is higher than the other two signals. So we can say that time domain wave multifractal is very sensitive for diagnosing the fault of the packing leakage in early stage.

The result is shown in TABLE.6 after getting the multifractal contrast of the first three groups in time-frequency domain.

Theoretically speaking, because the energy dissipation is the irreversible process[7], if there are some faults, the system must provide energy to overcome the fault. When the dissipation of energy

assumes the increasing tendency, it expresses as the increase in the dimensions of the fraction.

There is no big difference between the normal and the serious divulging in the time-domain multifractal by the TABLE. 5 and 6, but using the multifractal in time-frequency domain, we differentiate three kinds of situations from the quantification characteristic of the fractal figure, and the method has theoretical significance. So the multifractal in time-frequency domain can distinguish this kind of fault. For the fault as the filling leaking out, error diagnostics in the serious divulging were avoided, when we use the multifractal in time-frequency domain.

7. Conclusion

In this paper, vibration signal multifractal method was researched, multifractal on time-*frequency* domain was a energy fractal.. After the signal was changed to time-frequency domain by local wavelet decomposition, general dimension D_q would be calculated from the signal in time-frequency domain by least square method. The emulational analysis and practical application proved that fractal in the time-frequency domain fuses more information than in the time domain, and can exactly extract the signal's feature, .vibration signal multifractal method in the time-frenquency domain was feasible.in the practical application.

References

- [1] Shi Biqiang Shen yanhua, 2001, The fractal method of machinery fault diagnosis, Metallurgical Industry Press, Bei Jing.
- [2] Lv Jinhu Lu Junan Chen Shihua, 2001, Chaotic Time Series Analysis and Its Applications, Wu Han University Press, Wu Han.
- [3] Mandelbrot B B,The fractal Gemetry of nature,W H Freeman, San Francisco,1982
- [4] Ma Xiaojiang Yu Bo Zhang Zhixin,2000,Time-frequency analysis and application of Partial Wave Method Machine Tool & Hydraulics, VOL.3, pp.219-224.
- [5] Xu Yuxiu Yuan Peixin Yang Wenping, fractal and wavelet methods of complex mechanical fault diagnosis(M), Metallurgical Industry Press, Bei Jing, 2003.
- [6] Jiang Jiandong Wang Xiaosheng Qiu Liansheng, 1998, Chaos and Fractal Dynamics and It's Application in Machinery Faults Diagnosis, China Mechanical Engineering,Vol 7,pp.31-34.
- [7] Stane K, Matjaz P, Detecting chaos from a time series, Eur J.phys.26,pp.205-215.